IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

A NON-PROVISIONAL PATENT APPLICATION

FOR

DETECTING GENE EXPRESSION IN LIVE CELLS USING SHORT-LIVED REPORTERS WITH ENZYMATIC AMPLIFICATION

RELATED APPLICATIONS

The present application claims priority to and the benefit of US provisional application serial number 60/459,897, filed April 2, 2003.

FIELD OF THE INVENTION

The present invention relates to compositions and methods for detecting and analyzing gene expression events occurring in live cells.

BACKGROUND

One of the major challenges in the post-genomic era is to understand how genes are expressed and regulated. Gene expression can be tracked at the mRNA and protein level. Despite considerable progress in transcription and translational profiling with micorarray and mass spectrometry, methods that continuously monitor gene expression dynamics in live cells are in high demand. In addition, current microarray and mass spectrometry technologies cannot detect low copy number gene products, which often play a prominent role in sensing, signaling and gene regulation.

In recent years, tremendous progress has been made in the area of single-molecule detection in biological systems. It is fair to say that the single-molecule approach has changed the way many biological problems are addressed and interpreted. New insights derived from this approach are continuing to emerge. Although most of the single-molecule work has been carried out *in vitro*, single molecule experiments in living cells are beginning to appear. Indeed, gene expression in a single cell is a single molecule problem. In addition, the low copy numbers of

mRNA and proteins exhibit stochastic fluctuations similar to those seen in single-molecule experiments.

The use of reporter proteins have been employed to detect events of gene expression in cells. Typically, green fluorescent protein (GFP) and its derivatives are used as reporter proteins. The main advantage of GFPs is that they do not require exogenous substrate or cofactor. Most applications of GFPs have focused on mapping protein localization via fusion constructs. However, current GFPs are not suitable for following fast biological processes on the time-scale of minutes or less. This is primarily due to the fact that GFPs in the cellular environment have a long post-translational maturation time, which is required for the oxidation of the three residues forming the GFP fluorophore. Thus, a new GFP variant with faster maturation time is needed. However, even with such a variant, one GFP molecule only provides one fluorophore, thus it is only suitable for detecting translational product that expressed at high levels.

Currently, a need exists for a new reporting system that allows real-time detection of low copy number translational products in individual live cells. Moreover, there is a concomitant need for such a reporter system that employs compositions, which shorten the cellular lifetime of the reporter protein, thus allowing for following real-time biological processes while obtaining background-free measurements with high sensitivity.

SUMMARY OF THE INVENTION

The present invention pertains to compositions and methods for detecting and analyzing gene expression events occurring in live cells. In one aspect, the present invention pertains to a short-lived reporter with enzymatic amplification. The reporters of the present invention have relatively short maturation time and a short cellular lifetime which can be exploited to detect transient events of gene expression in live cells.

In one embodiment of the present invention, compositions and methods for employing one or more reporters having a short maturation time and a short cellular

lifetime to detect transient events of gene expression in individual living cells with high sensitivity and high time resolution are described. Also, described herein is a reporter gene system employing a reporter, for example, β-galactosidase (β-gal). In one aspect of this embodiment, the reporter is manipulated in such a manner so as to decrease its cellular lifetime.

In this aspect, the so-called N-end rule to shorten the cellular lifetime of β-gal is utilized. The N-end rule states that the cellular lifetime of a protein is related to its N-terminal amino acid residue. This rule applies to all organisms ranging from bacteria to mammals. In *E. coli*, changing the N-terminal amino acid from the natural methionine to leucine, arginine, lysine, phenylalanine, tryptophan or tyrosine shortens the protein half-life to a few minutes. Since all newly translated proteins have methionine at the N-terminus (the translation start codon encodes for methionine), the ubiquitin (ub) fusion technique is used to introduce a lifetime-shortening amino acid (*e.g.*, leucine or arginine) in place of the methionine at the N-terminus of, for example, β-gal to generate Ub-Leu-β-gal or Ub-Arg-β-gal. After this reporter protein is expressed, the ubiquitin will be cleaved by an ubiquitin-specific protease, thus exposing the leucine or arginine residue and targeting the protein for the proteolytic pathways. In addition to the N-end rule, other means of modifying β-gal's cellular lifetime are also employed, such as N-terminal and C-terminal signal peptides fusions.

In another embodiment, live-cell microarrays are described. In this embodiment, multiple libraries of cells are prepared each differing in at least one genotypic property (*i.e.*, the genotype of each cell is different, for example, the reporter gene is inserted at a different position on the chromosome, thereby tagging an operon or a gene). In one aspect, a live-cell microarray is comprised of two libraries. One library comprises cells each of which has a promoterless *lacZ* gene encoding for a short-lived β-gal with its own ribosome binding site that is inserted into one promoter controlled region in the host cell's genome. The second library comprises the same elements except that a gene encoding for a short-lived yellow fluorescent protein YFP (Venus-ssrA) replaces the gene encoding for a short-lived β-gal in the first library.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 (a) are chemical structures of 9H-(1,3-dichloro-9,9-dimethylacridin-2-one-7-yl) β-D-galactopyranoside (DDAO-gal) and its fluorescent product DDAO after hydrolysis by β-galactosidase, and (b) is a graphical representation of the absorption and emission spectra of DDAO; (c) are chemical structures of resorufin-glucopyranoside (resorufin-glu) and its fluorescent product resorufin after hydrolysis by β-glucosidase and (d) is a graphical representation of the absorption and emission spectra of resorufin;

FIG. 2 (a) shows the location of the gene coding for Ub-Arg-β -gal in the lac operon and (b) depicts the nucleic acids and amino acids sequences of Ub-Arg-β-gal. Only the sequences of ubiquitin (light-shaded), the arginine residue immediately after ubiquitin, and the linker peptide (unshaded) between ubiquitin and the beginning of β-gal (dark-shaded) are shown. Please note that the β-gal in this construct lacks its first twenty two amino acids;

FIG. 3 (a) is a graph measuring the hydrolysis of DDAO-gal in the presence of enzyme products from different gene constructs, and (b) are the amino acid (top) and nucleotide sequence (bottom) for each of the different construct; Please note that only the sequences that differ in these constructs (N-terminus of the *lacZ* gene) are shown in (b);

FIG. 4 is a graph showing the DDAO fluorescence generated from the hydrolysis of DDAO-gal by wild type $lacZ^+$ cells (dark) but not by the $lacZ^-$ cells (light);

FIG. 5 (a) depicts the sequence junction of *lacZ* deletion, wherein the sequence is from the EcorV site of the *lacI* gene to NspI site of the *lacY* gene, and (b) is the amino acid sequence and nucleic acid sequence wherein the numbering of the nucleotides is according to the first base of the *lacI* gene, the *lacZ* gene is replaced by

lacY gene from the ATG starting codon, the amino acids sequences are shown on top of the DNA sequence panel;

FIG. 6 is a fluorescence image of *E. coli* Cells. The signal is from DDAO generated by the basal level expression of unmodified β-gal;

FIG. 7 (a) is the fluorescence images observed on single *E.coli* cells with a gene coding for a short-lived Ub-Arg-β-gal incorporated on chromosome. The signal is generated by the basal level expression of β-gal, For cell 1, only thirteen fluorescence images of cell 1 are shown in fifteen minute intervals for simplicity reasons. For cell 2, the fluorescence images are shown in five minute intervals. (b) is a fluorescence measruement of the production and degradation of β-gal in one singe *E.coli* cell under TIR fluorescence microscope;

FIG. 8 (a) depicts the sequence for the short-lived YFP: Venus-ssrA construct on plasmid pVS5, and (b) is the amino acid sequence and nucleic acid sequence wherein the sequence is from the first base of the *yfp* gene and to the end of the *yfp* gene with the addition of 33 bases coding for the ssrA tag;

FIG. 9 is a graph showing the resorufin fluorescence generated from the hydrolysis of resorufin-glu by *E.coli* cells expressing β-glucosidase (*bgl*B$^+$, light) but not by the *bgl*B$^-$ cells (dark);

FIG. 10 is a schematic drawing of the construction of a *lacZ* library by Tn5 mediated transposition;

FIG. 11 is a schematic drawing of the constructing a *lacZ* and *yfp* library;

FIG. 12 is a flow chart showing an automated process for the construction of libraries and the fabrication of the cell array;

FIG. 13(a) is the plasmid map for pBBR1MCS-5.1, and (b) is the nucleotide sequence coding for the short-lived b-gal for the plasmid depicted in (a). Please note that only the sequence at the N-terminus of the *ub-leu-lacZ* gene is shown;

FIG. 14 (a) is a fluroescence image of *Shewanella oneideinis* cells expressing β-gal from the *lacZ*$^+$ plasmid pBBR1MCS5.1, and (b) is a graph showing the DDAO fluorescence generated by the hydrolysis of DDAO-gal under various conditions;

FIG. 15 (a) depicts the nucleotide sequence junction of *ub-leu-lacZ* gene in *Saccharomyce cerevisiae* and (b) is the amino acid sequence and nucleic acid sequence for the junction of the *ub-leu-lacZ* construct on centromeric plasmid transformed into *Saccharomyce cerevisiae* cell;

FIG. 16 represents DDAO fluorescence generated from the hydrolysis of DDAO-gal by wild type *lacZ*$^+$ cells (dark) but not by the *lacZ*$^-$ cells (light) in *Saccharomyce cerevisiae*;

FIG. 17 is a fluorescence image of *S. cerevisiae* cells containing unmodified β-gal; and

FIG. 18 is the fluorescence signal bursts observed on a single *S. cerevisiae* cell with a short-lived β-gal expressed from a centromeric plasmid.

DETAILED DESCRIPTION

The present invention pertains to compositions and methods for detecting and analyzing gene expression events occurring in individual living cells. In particular, the present invention pertains to short-lived reporters with enzymatic amplification. These reporters of the present invention have relatively short maturation time and a short cellular lifetimes which can be exploited to detect transient events of gene expression in live cells.

Tremendous progress has been made to track gene expression at the mRNA level by DNA arrays and at the protein level by mass spectrometry. Although current

DNA microarray and mass spectrometry technologies have started to address compelling biological problems at a genome-wide scale, they suffer from a few disadvantages: (1) they cannot continuously monitor temporal evolution of expression - multiple samples have to be taken in order to evaluate the response to a stimulus or an environmental change; (2) they cannot follow fast gene expression processes on the time scale of minutes. The low time resolutions prevent studies of transient gene expression processes, for example, those involved in cell division; (3) they are not sensitive to low copy number gene products, which often play a prominent role in cellular sensing, signaling and gene regulation; and (4) they can only provide averaged results of large populations of cells rather than behaviors of individual cells: transient and stochastic gene expression events are often masked in the population measurements.

In one embodiment of the present invention, a method for employing one or more reporters having a short maturation time and a short cellular lifetime to detect transient events of gene expression in live cells with high sensitivity and a fast time resolution is described.

In one aspect, a reporting system for monitoring real-time gene expression events in a living cell is disclosed. This reporting system comprises an illuminogenic substrate, wherein said substrate is permeable to said cell. The system also comprises at least one reporter protein, wherein said reporter protein facilitates the conversion of said illuminogenic substrate into an illuminescent molecule, and wherein said reporter protein has a short cellular life time.

In this aspect, the cell can be a prokaryote or eurokaryote. The illuminogenic substrate can be any substrate that when acted upon by, for example, hydrolysis, will generate an illuminescent product which emits photons. For example, the substrate can be a fluorogenic substrate that when acted upon will generate a fluorescent product that emits fluorescence. Chemiluminescence substrates can also be used in the present invention. The term illuminogenic is also meant to cover absorption in addition to photon emission, for example, chromogenic substrates.

The present embodiment is designed to capitalize on the recent advances in sensitive fluorescence microscopy. In the past years, tremendous progress has been made in fluorescence imaging of single-molecules, even in living cells. See, for example, Sako, Y. and T. Uyemura, Total Internal Reflection Fluorescence Microscopy for Single-molecule Imaging in Living Cells. Cell Struct Funct, 2002. 27(5): p. 357-65; Sako, Y., S. Minoghchi, and T. Yanagida, Single-molecule imaging of EGFR signalling on the surface of living cells. Nat Cell Biol, 2000. 2(3): p. 168-72; Seisenberger, G., *et al.*, Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. Science, 2001. **294**(5548): p. 1929-32; and the entire teachings of which are incorporated herein by reference. State-of-the-art microscopes are more than capable of imaging single or multiple numbers of gene products of a single gene, if not single fluorophores in a live cell.

A popular approach for real-time observation of gene expression in live cells is

the use of green fluorescent protein (GFP) and derivatives thereof as reporter proteins. See, for example, Bongaerts, R.J., *et al.*, Green fluorescent protein as a marker for conditional gene expression in bacterial cells. Methods Enzymol, 2002. **358**: p. 43-66; Tsien, R.Y., The green fluorescent protein. Annu Rev Biochem, 1998. **67**: p. 509-44; and Chalfie, M., *et al.*, Green fluorescent protein as a marker for gene expression. Science, 1994. **263**(5148): p. 802-5, the entire teachings of which are incorporated herein by reference. The main advantage of GFPs is that they do not require an exogenous substance or cofactor. Most applications of GFPs have focused on mapping protein localization via fusion constructs. However, GFPs are not suitable for following faster biological processes on the time-scale of minutes or less. This is due to the fact that GFPs in cellular environments have a long post-translational maturation time (Perozzo, M.A., *et al.*, J Biol Chem, 1988. **263**(16): p. 7713-6, and Heim, R., D.C. Prasher, and R.Y. Tsien, Proc Natl Acad Sci U S A, 1994. **91**(26): p. 12501-4, the entire teachings of which are incorporated herein by reference), which is required for the oxidation of the three residues forming the GFP fluorophore.

As a reporter gene, one GFP molecule only provides for one fluorophore, thus high sensitivity detection is required for low copy numbers. Described herein is a

reporter gene system that circumvents these difficulties. To illustrate this new system, β-galactosidase ("β-gal") is used, however, it should be obvious to those skilled in the art that other reporter genes can equally be the subject of the present invention such as β-glucosidase.

Other enzyme-substrate systems that can be employed include, but are not limited to, the following: (a) enzyme: β-galactosidase, substrates: DDAO-galactopyranoside, Resorufin- galactopyranoside; (b) enzyme: β-gluocosidase, Substrates: Resorufin-glucopyranoside, DDAO-glucopyranoside; (c) enzyme: β-lactamase, substrate: CCF2 (see, Zlokarnik *et al.*, Science, 1998, 279(5347), 84-88, and CR2/AM (Gao *et al.*, J.Am.Chem. Soc, 2003, 125, 11146-11147, the entire teachings of which are incorporated herein by reference.) It should be understood that other enzyme activities similar to those just listed are also encompassed within the present invention. Additionally, modified proteins having identical or similar enzymatic activities are also encompassed within the present invention. For example, proteins that have between 45% to 65% structural homology (and similar enzymatic activity) with the enzymes described herein are within the scope of the invention. (Unless otherwise stated, the terms protein and peptide can be used interchangeably herein.) Proteins having between 65% to 75% structural homology (and similar enzymatic activity) with the enzymes mentioned above are within the scope of the invention. Proteins having between 75% to 85% structural homology (and similar enzymatic activity) with the enzymes described herein are within the scope of the instant invention. Protein having between 85% to 100% structural homology (and similar enzymatic activity) with the enzymes described herein are within the scope of the present invention.

β-gal is a well-studied reporter (encoded by the *lacZ* gene of *E. coli*) and has a relatively short maturation time and fast enzymatic hydrolysis rate of fluorogenic substrates. DDAO-gal (from Molecular Probes) is a good fluorogenic substrate for assaying β-gal activity *in vivo* (FIG. 1a). DDAO's emission maximum is at 660 nm (FIG. 1b), having little overlap with autofluorescence of the cell, making it highly suitable for live cell studies. Because one copy of the enzyme (β-gal) generates approximately one thousand fluorescent DDAOs per second, the fluorescent signal is

amplified by the enzymatic reaction, making it possible to detect low copy numbers of β-gal. Without induction, there are about 10 β-gal per *E.coli* cell (Sambrook, J. and D. Russell, *Molecular Cloning*. 3rd ed. Vol. 3. 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. 15.57, the entire teaching of which is incorporated herein by reference), providing a good model system for detecting genes that expressed at low copy numbers. (It should be noted that other substrates can also be employed such as resorufin-glu, whose hydrolyzed product resorufin has am maximum absorption at 571 nm, and emission at 585 nm.)

On the chromosome of *E.coli*, β-gal expression is stochastic. Without inducers, a *lac* repressor binds tightly to a DNA sequence known as the *lac* operator. When it occasionally falls off the operator sequence of the chromosome, one or more copies of mRNA followed by a few copies of β-gal are produced through transcription and translation. DDAO-gal can be used to observe this stochastic event of β-gal expression.

However, in *E. coli*, the lifetime of β-gal is longer than 10 hours, Tobias, J.W., *et al.*, Science, 1991. **254**(5036): p. 1374-7, the entire teaching of which is incorporated herein by reference. This presents a general obstacle to follow dynamic processes. A long-lived reporter protein leaves a constant background that prevents the detection of small and transient variations. A solution to this problem, which is the subject of this invention, is to shorten the cellular lifetime of reporter proteins. The reporter proteins are degraded shortly after they are expressed, generating a background free condition for sensitive detection. This provides a general approach for visualizing individual gene expression events in realtime, as these events can be observed as discrete fluorescence bursts.

In one aspect of the invention, the so-called N-end rule to shorten the half-life of β-gal is used, see, Tobias, J.W., *et al.*, Science, 1991. **254**(5036): p. 1374-7, the entire teachings of which are incorporated herein by reference. The N-end rule states that the cellular half-life of a protein is related to its N-terminal amino acid residue. This rule applies to all organisms ranging from bacteria to mammals. In *E. coli*, changing the N-terminal amino acid from the natural methionine to leucine, arginine,

lysine, phenylalanine, tryptophan or tyrosine shortens the protein half-life to about two minutes. Since all newly translated proteins have methionine at the N-terminus (the translation start codon encodes for methionine), the ubiquitin fusion technique is used to introduce a lifetime-shortening amino acid (*e.g.*, leucine or arginine) in place of the methionine at the N-terminus of β-gal to generate Ub-Leu-β-gal or Ub-Arg-β-gal, Seisenberger, G., *et al.*, Science, 2001. 294(5548): p. 1929-32, the entire teaching of which is incorporated herein by reference. After this reporter protein is expressed, the ubquitin will be cleaved by an ubiquitin-specific protease, thus exposing the leucine or arginine residue and targeting the protein for the proteolytic pathways.

Using an *ub-arg-lacZ* reporter gene (coding for Ub-Arg-β-gal) on the chromosome of *E.coli*, it has been demonstrated that an *in vivo* half-life of about two minutes in *E. coli* can be obtained. Figure 2a depicts the *ub-arg-lacZ* gene in a chromosomal positioning alignment. Figure 2b provides the nucleotide sequence [SEQ ID NO. 1] and amino sequence [SEQ ID NO 2].

To generate this *ub-arg-lacZ* reporter gene, a pair of PCR primers (5' GATG GATCCGTCGTTGCTGATTGGCGTTG 3', [SEQ ID NO. 3] and 5' GATGGATCC CGCAGGCTTCTGCTTCAATC 3', [SEQ ID NO. 4]) were used to amplify a 2000 bp fragment containing partial *lacI*, complete *lac* operon regulation region (the sequence between the end of the *lac*I gene and the beginning of the *lacZ* gene) and partial *lacZ* gene from the *E.coli* strain k12 chromosome DNA. This fragment was then digested by BamHI, and ligated into a BamHI digested plasmid pBR322 (New England Biolabs) to create plasmid pBR322-IZ using standard cloning protocols Sambrook and Russell, Molecular Cloning, 3rd Ed, CSHL press. Another pair of inverse PCR primers (5' CATAGCTGTT TCCTGTGTGAAATTGTTATCCGC 3', [SEQ ID NO.5] and 5' GGTGCCGGAA AGCTGGCTGGAG 3', [SEQ ID NO. 6]) was used to open this newly constructed pBR322-IZ at the 3' position of the starting codon ATG of the *lacZ* gene. A third pair of PCR primers (5' CAGATTTTCGTCAAGACTTT GACC3', [SEQ ID NO. 7] and 5' GCTTCTGGTGCCGGAAAC 3', [SEQ ID NO. 8]) were used to amplify the ubiquitin gene, the arginine residue (codon AGG) immediately after the C-terminal glycine of ubiquitin and the linker sequence between ubiquitin and *lacZ* from plasmid

pUB23-arg (gift from Professor Daniel Finley, Harvard Medical School). This DNA fragment was ligated into the inverse PCR-opened pBR322-IZ and the orientation of the ubiquitin relative to the *lacZ* gene was verified by DNA sequencing. Next, the replacement of the wild type *lacZ* gene on the *E. coli* chromosome was achieved by homologous recombination using a gene replacement vector pKO3, see, Link, *et al.*, J Bacteriol, 1997. 179(20): p. 6228-37, the entire teaching of which is incorporated herein by reference. The final resulting construct on the chromosome is depicted in FIG. 2.

In addition to this Ub-Arg -β-gal construct, a repertoire of short-lived β-gals with different cellular lifetimes were constructed. In one group (N-end rule), the linker sequence lying between ubiquitin and *lacZ*, referred to as "eK" sequence, was varied. See FIG. 3(a) k12-e1 [SEQ ID NOS. 9, 10], k12-e2 [SEQ ID NOS. 11, 12] and k12-e3a [SEQ ID NOS. 13, 14] (where the top row in the sequence identification represents the amino acid sequence and the botton two rows represent nucleotide sequence), where the light-shaded sequence is ubiquitin (Tobias et al., Science, 1991, 254, 1374, the entire teaching of which is incorporated herein by reference), the unshaded sequences are the linker sequence between ubiquitin and *lacZ*, of which and the length and amino acids compositions are altered, and the dark-shaded sequence is the beginning of the *lacZ* gene without the first twenty two amino acids. It has been demonstrated that in addition to the leucine residue immediately after ubiquitin, the linker sequence has a profound impact on the cellular lifetime of β-gal. The hydrophobicity of the amino acids composition and the length (or disordered structure) contributes greatly to the overall recognition and delivery of β-gal to downstream proteases.

In another group, N-terminal signal peptides derived from naturally short-lived proteins are fused to the beginning of β-gal to shorten its cellular lifetime. The two strains k12-n3 [SEQ ID NOS. 15, 16] and k12-n5 [SEQ ID NOS. 17, 18] (where the top row in the sequence identification represents the amino acid sequence and the bottom two rows represent the nucleotide sequence) depicted in FIG. 3(a) belong to this group (N-terminal modification). In the sequence panel of k12-n3 and k2-n5, the light-shaded sequences are signal peptides taken from the published work of Flynn *et*

al., (Flynn, JM. *et al.*, Molecular Cell, 2003, 11, 671-683, the entire teaching of which is incorporated herein by reference), and the dark-shaded sequence is the beginning of the *lacZ* gene without the first methionine.

FIG 3(b) illustrates the different cellular lifetimes of these modified β-gals expressed from *E.coli* chromosome, as indicated by the different DDAO-gal hydrolysis rates. The measurements were done using a fluorometer, in which DDAO-gal at a final concentration of 100 μM was added to *E. coli* cells grown to middle log phase in M9 minimal media. The fluorescence of the hydrolyzed product, DDAO, was monitored over time at 660 nm with excitation at 638 nm. The hydrolysis rate was then calculated by measuring the slope of the fluorescence increase over time. As a reference, the DDAO-gal hydrolysis rate by the wild type k12 strain is also shown.

To illustrate the use of β-gal as a reporter gene, investigators chose *E.coli* as a test organism. A plasmid encoding for ampicillin resistance gene β-lactamase was transformed into *E.col* strains. The presence of the β-lactamase allows the usage of the antibiotic ampicillin, which not only keeps the contamination of other bacteria minimal, but also increases the permeability of the *E.coli* cell wall to the fluorogenic substrate DDAO-gal. The mechanism of the increased cell wall permeability is very likely due to the known fact that ampicillin inhibits cell wall synthesis. All the strains described in this invention contain such an ampicillin-encoding plasmid. Figure 4 shows the measurements of the DDAO fluorescence signal generated by the hydrolysis of DDAO-gal in the wild type *E.coli* cells. The measurements were done using a fluorometer under the same conditions as described in FIG 3(b). A fluorescence signal increase can be observed immediately upon the addition of the substrate, demonstrating that DDAO-gal can permeate through cell wall and inner membrane of *E. coli*.

In contrast, as a control experiment, a negligible rate of DDAO-gal hydrolysis was observed in a *lacZ* deficient strain (*lacZ*$^-$) of *E. coli* (FIG. 5 depicts both the amino acid sequence [SEQ ID NO. 19] and the DNA sequence [SEQ ID NO. 20] around the region where *lacZ* is deleted from chromosme.) that is primarily due to

autohydrolysis. This experiment demonstrates that the hydrolysis of DDAO-gal is specific to the presence of β-gal.

The microscopy experiment was performed using a through-lens total internal reflection (TIR) microscope from Olympus and an intensified CCD camera from Roper Scientific. The total internal reflection excitation allows detection of only a thin layer
(< 400 nm) above the cover slip and effectively suppresses the fluorescene background of the medium. The excitation light was set at 638 nm, wherein autofluorescence of the *E. coli* cell is negligible. This detection system assures the highest sensitivity available. The sample chamber (Bioptech) was maintained at 37°C with M9 minimal medium perfusing through the chamber. *E. coli* cells were pushed down on the glass coverslip by a droplet of agarose gel.

As shown in FIG. 6, a strong DDAO fluorescence signal from individual *E. coli* cells with wild type β-gal (long lifetime about 10 hours) was detected. This was done at the basal level, *i.e.*, the *lacZ* gene is not induced. DDAO can diffuse out or be expelled by the cell. Once it leaves the cell, DDAO quickly diffuses out from the probe volume. A steady signal was observed. In contrast, as shown in FIG. 7, when the gene encoding for a short-lived Ub-Arg-β-gal (see FIG 2 for sequence) replaced the wild type *lacZ* gene encoding for the long-lived β-gal on the chromosome, single fluorescence bursts corresponding to stochastic expression of the *lacZ* gene were observed in real time in single *E.coli* cell (see FIG. 7(a) for the fluorescence images of *E.coli* cells). Each burst is triggered by the dissociation of the *lac* repressor from the *lac* operator on the *E.coli* chromosome. The fluorescence off time corresponds to the time required for the repressor to dissociate from the operator sequence, while the fluorescence on time corresponds to the time required for the degradation of β-gal. Moreover, as shown in FIG. 7(b), the time trace of the fluorescence bursts exhibits quantized levels corresponding to β-gal molecules generated and degraded one molecule at a time. This demonstrates the signal molecule's sensitivity of this reporting system.

In this embodiment, in order to detect genes with higher expression levels, a short-lived version of a yellow fluorescent protein (YFP) variant, Venus, (Venus-ssrA) is employed. Extensive randomized and directed mutagenesis efforts have produced various GFP and YFP derivatives with faster maturation time than the wild type GFP (30-90 minutes), (Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44, the entire teaching of which is incorporated herein by reference), thereby enabling the use of GFPs and YFPs as reporters for transient dynamic changes. One of the more promising YFP variants is "Venus," which matures in ~ 3 minutes *in vitro*, see, Nagai, T., *et al.*, Nat Biotechnol, 2002. **20**(1): p. 87-90, the entire teaching of which is incorporated herein by reference. Like other YFPs, the matured Venus is stable in the cell with a lifetime of ~24 hours, see, Li, X., *et al.*, J Biol Chem, 1998. **273**(52): p. 34970-5, the entire teaching of which is incorporated herein by reference.

One aspect in particular pertains to a short-lived Venus variant by creating a Venus-ssrA construct. In this construct, the ssrA peptide tag sequence (AANDENYAKAAA, [SEQ ID NO. 21]) was encoded at the DNA level as a C-terminal fusion to Venus. Normally, a bacterial cell uses a ssrA sequence to flag a protein as the result of a prematurely terminated translation (see Kenneth C. Keiler, Patrick R. H. Waller, Robert T. Sauer, Science, 1996, **271**, 990-993). Tagging Venus with ssrA tag recruits cellular protein degradation machinery and greatly reduces the cellular lifetime of Venus from more than 24 hours to less than 30 minutes. It is straightforward to extend this strategy to other GFP variants for construction of other GFP based short-lived reporter proteins.

FIG 8(a) illustrates plasmid pVS5 which encodes the Venus-ssrA gene. FIG 8 (b) shows the nucleotide [SEQ ID NO. 22] and amino acid sequences [SEQ ID NO. 23] of the Venus-ssrA gene. The first amino acid shown in the figure is the first amino acid of Venus. To generate the venus-ssrA reporter gene, a pair of PCR primers (5' CACCAGC AAGGGCGAGGAGCTGTTC-3' [SEQ ID NO. 24] and 5' TTCTTAGGCGGCTAAGG CGTAGTTCTCGTCGTTGGCGGCCTTGTACAGCTCGTCCATGC-3' [SEQ ID NO. 25]) were used to amplify the Venus gene from a plasmid pCS2/venus (Nagai T,

Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A. Nat Biotechnol. 20(1):87-90) and add the ssrA sequence at the 3' end of the Venus gene. The resulting PCR fragment was then ligated into pBAD202/TOPO vector (Invitrogen Inc.) to generate plasmid pVS5.

Again, this general strategy allows for highly sensitive detection of dynamic processes in living cells free from the complication of large fluorescence background associated with protein accumulation.

Short-lived β-gal and short-lived YFP are complimentary to each other. Short-lived β-gal can be used to detect genes that are expressed at low copy numbers because of the enzymatic amplification. Short-lived YFP provides a linear response to high-level gene expression. Real-time analysis of short-lived-YFP-incorporated cells typically work under aerobic conditions, while short-lived β-gal incorporated cells typically work under both aerobic and anaerobic conditions. The combination of the two reporter proteins will cover a broad range of intracellular gene expression levels and applicable organisms

In another embodiment, compositions and methods are described for live-cell microarrays. In this embodiment, multiple libraries of cells each differing in at least one genotypic property are prepared. In one aspect, a live-cell microarray is comprised of two libraries. One library comprises cells each of which has a promoterless *lacZ* gene encoding for a short-lived β-gal with its own ribosome binding site that is operatively linked to one promoter controlled region in the host cell's genome. The second library comprises the same elements except that a gene encoding for a short-lived YFP (Venus-ssrA) replaces a gene encoding for a short-lived β-gal.

The construction of the libraries can be accomplished by random insertion mediated by transposition or by homologous recombination. DNA sequencing around the insertion of the cells in the library will allow a practitioner to identify the position of the insertion with respect to the genome. In one particular aspect, a 75 x 75 element array is sufficient to contain a library with one insertion per gene for a

genome has approximately 4000 genes (*E.coli* has about 4000 genes). (It should be noted that one skilled in the art will appreciate that various other arrays can be employed.) In another particular aspect, instead of inserting each reporter per gene, the reporter is operatively linked per operon. The size of the array can be smaller if only one insertion is allowed per promoter-controlled region.

Two sets of live-cell microarrays are made from the two libraries of cells with, for example, liquid handling robots preparing the cells on a substrate such as a glass slide with a micro droplet of agarose containing growth media on top of the cells in order to immobilize the cells for ease of measurement, storage and transportation.

Examining the microarrays under a fluorescence microscope, one can study gene expression responses to stimuli and/or environmental changes. For example, parallel movies of all elements of the microarrays can be recorded and vast amounts of data can be compiled and analyzed. The microarrays provide first-of-a-kind genome-wide gene expression profiling and massive kinetics data with high sensitivity and time resolution in living cells

The advantages of employing live-cell microarrays can be summarized as follows:
(1) real-time and parallel observations; (2) high throughput system-wide profiling; (3) quantitative analyses of gene expression levels; (4) background free measurements due to short cellular lifetime of the reporter proteins; (5) high sensitivity for low copy number genes due to enzymatic amplification; (6) single cell sensitivity enabling observation of stochastic events; (7) high time resolution (minute) allowing observations of transient behaviors; (8) broad dynamics range afforded by the combination of two reporter genes; (9) ease and low cost in studying the microarrays with commercially available fluorescence microscopes in non-specialized laboratories; and (10) low cost in microarray replication for distribution to the scientific community.

In one aspect, one cell per element of the microarray (*e.g.*, 100μm x 100μm) can be effectuated. In order to obtain reliable statistics, however, one may wish to

place a larger number of cells (10-100) per array element. In addition, high sensitivity makes it possible to observe the behavior of single bacterial cells in a microbial community. Not only can one detect common trends in expression profiles, a practitioner can also observe how gene expression in one cell affects its neighbors, allowing an investigator to pinpoint cooperative effects among cells. Finally, with the background rejection advantage of confocal or total internal reflection microscopy, one has both the high sensitivity to detect low-level expression events and the ability to penetrate multiple layer of biofilm.

It is important to stress that the present invention possesses significant sensitivity for detecting a single copy of reporter proteins in single cells, as exemplified in the Example section (see below). This allows stochastic events of gene expression of low copy number genes to be observed. Stochasticity of gene expression has attracted many experimental and theoretical efforts recently. Combined with the live-cell arrays and short-lived reporter proteins, the highly sensitive measurements of gene expression provide unprecedented information on the working of the genetic network of a genome.

Systematic analyses of the gene expression patterns and their temporal evolution are expected to provide detailed information and generate new insights into function and control of gene expression processes.

In one embodiment of the present invention, cell sorting is facilitated by the compositions described herein. In this embodiment, an illuminogenic substrate, such a fluorescence substrate is introduced to a cell or population of cells, wherein the substrate enters the cells. A nucleotide sequence encoding a reporter protein of the instant invention is also introduced to the cells and is operatively linked within the cell's genome. For instance, the reporter gene (*i.e.*, the nucleotide sequence encoding for the reporter protein) can be operatively linked to a predetermined host gene.

As described above, the reporter protein comprises enzymatic activity such that when it is expressed within a host cell it can facilitate the conversion of the illuminogenic substrate to an illuminesence molecule. With this system in place, a

practitioner can examine various perturbations made upon the cell or cell population and determine if a particular perturbation or set of perturbations trigger the translation of a particular protein. If a particular gene, which is operatively linked to a reporter gene, is expressed upon a perturbation(s) to the cell or any of its components, then an illuminogenic signal will be emitted.

Cells emitting a particular signal can then be separated from cells not emitting such a signal. For example, conventional fluorescence cell sorters are available and can be employed in this embodiment.

Agents used to perturb a cell can include, but not limited to, pharmaceutical agents, including test agents, pesticides, chemical agents both gaseous and in liquid form, hormones, metabolites, toxins, pheromones, and alike.

To facilitate the understanding of the present invention, a number of terms and phrases are defined below:

As used herein, the term "nucleotide" is used to include polymeric forms of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Nucleotides can have any three-dimensional structure, and can perform any function, known or unknown. The following are non-limiting examples of nucleotides: a gene or gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant nucleotides, branched nucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A nucleotide can comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure can be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A nucleotide may be further modified after polymerization, such as by conjugation with a labeling component. The term also includes both double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of this invention that is a nucleotide encompasses both the double-stranded form and

each of two complementary single-stranded forms known or predicted to make up the double-stranded form.

A nucleotide is composed of a specific sequence of four nucleotide bases: adenine (A); cytosine (C); guanine (G); thymine (T); and uracil (U) for thymine when the polynucleotide is RNA. Thus, the term "nucleotide sequence" is the alphabetical representation of a nucleotide molecule. This alphabetical representation can be inputted into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

A "gene" includes a nucleotide containing at least one open reading frame that is capable of encoding a particular polypeptide or protein after being transcribed and translated. Any of the nucleotide sequences described herein may be used to identify larger fragments or full-length coding sequences of the gene with which they are associated. Methods of isolating larger fragment sequences are known to those of skill in the art, some of which are described herein.

A "gene product" includes an amino acid, *e.g.*, peptide or polypeptide, generated when a gene is transcribed and then translated.

A "primer" includes a short nucleotide, generally with a free 3'-OH group that binds to a target or "template" present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a nucleotide complementary to the target. A "polymerase chain reaction" ("PCR") is a reaction in which replicate copies are made of a target polynucleotide using a "pair of primers" or "set of primers" consisting of "upstream" and a "downstream" primer, and a catalyst of polymerization, such as a DNA polymerase, typically a thermally-stable polymerase enzyme. Methods for PCR are well known in the art, and are taught, for example, in MacPherson *et al.*, IRL Press at Oxford University Press (1991). All processes of producing replicate copies of a nucleotide, such as PCR or gene cloning, are collectively referred to herein as "replication". A primer can also be used as a probe in hybridization reactions, such as Southern or Northern blot analyses (*see*, for example, Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A*

Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

The term "cDNAs" includes complementary DNA, that is mRNA molecules present in a cell or organism made into cDNA with an enzyme such as reverse transcriptase. A "cDNA library" includes a collection of mRNA molecules present in a cell or organism, converted into cDNA molecules with the enzyme reverse transcriptase, then inserted into "vectors" (other DNA molecules that can continue to replicate after addition of foreign DNA). Exemplary vectors for libraries include bacteriophage, viruses that infect bacteria, *e.g.*, λ phage. The library can then be probed for the specific cDNA (and thus mRNA) of interest.

A "delivery vehicle" includes a molecule that is capable of inserting one or more nucleotides into a host cell. Examples of delivery vehicles are liposomes, biocompatible polymers, including natural polymers and synthetic polymers; lipoproteins; polypeptides; polysaccharides; lipopolysaccharides; artificial viral envelopes; metal particles; and bacteria, viruses and viral vectors, such as baculovirus, adenovirus, and retrovirus, bacteriophage, cosmid, plasmid, fungal vector and other recombination vehicles typically used in the art which have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts. The delivery vehicles may be used for replication of the inserted nucleotide, gene therapy as well as for simply polypeptide and protein expression.

A "vector" includes a self-replicating nucleic acid molecule that transfers an inserted polynucleotide into and/or between host cells. The term is intended to include vectors that function primarily for insertion of a nucleic acid molecule into a cell, replication vectors that function primarily for the replication of nucleic acid and expression vectors that function for transcription and/or translation of the DNA or RNA. Also intended are vectors that provide more than one of the above function.

A "host cell" is intended to include any individual cell or cell culture that can be or has been a recipient for vectors or for the incorporation of exogenous nucleic acid molecules, nucleotides and/or proteins. It also is intended to include progeny of

a single cell. The progeny may not necessarily be completely identical (in morphology or in genomic or total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. The cells may be prokaryotic, include but are not limited to bacterial cells.

The term "genetically modified" includes a cell containing and/or expressing a foreign gene or nucleic acid sequence that in turn modifies the genotype or phenotype of the cell or its progeny. This term includes any addition, deletion, or disruption to a cell's endogenous nucleotides.

As used herein, "expression" includes the process by which nucleotides are transcribed into mRNA and translated into peptides, polypeptides, or proteins. If the nucleotide is derived from genomic DNA, expression may include splicing of the mRNA, if an appropriate eukaryotic host is selected. Regulatory elements required for expression include promoter sequences to bind RNA polymerase and transcription initiation sequences for ribosome binding. For example, a bacterial expression vector includes a promoter such as the *lac* promoter and for transcription initiation the Shine-Dalgarno sequence and the start codon AUG (Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Similarly, a eukaryotic expression vector includes a heterologous or homologous promoter for RNA polymerase II, a downstream polyadenylation signal, the start codon AUG, and a termination codon for detachment of the ribosome. Such vectors can be obtained commercially or assembled by the sequences described in methods well known in the art, for example, the methods described below for constructing vectors in general.

"Differentially expressed", as applied to a gene, includes the differential production of mRNA transcribed from a gene or a protein product encoded by the gene. A differentially expressed gene may be overexpressed or underexpressed as compared to the expression level of a normal or control cell. In one aspect, it includes a differential that is 2.5 times, preferably 5 times or preferably 10 times higher or lower than the expression level detected in a control sample. The term "differentially

expressed" also includes nucleotide sequences in a cell or tissue which are expressed where silent in a control cell or not expressed where expressed in a control cell.

The term "peptide" includes a compound of two or more subunit amino acids, amino acid analogs, or peptidomimetics. The subunits may be linked by peptide bonds. In another embodiment, the subunit may be linked by other bonds, *e.g.*, ester, ether, etc. As used herein the term "amino acid" includes either natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. A peptide of three or more amino acids is commonly referred to as an oligopeptide. Peptide chains of greater than three or more amino acids are referred to as a polypeptide or a protein.

"Hybridization" includes a reaction in which one or more nucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a nucleotide by a ribozyme.

Hybridization reactions can be performed under conditions of different "stringency." The stringency of a hybridization reaction includes the difficulty with which any two nucleic acid molecules will hybridize to one another. Under stringent conditions, nucleic acid molecules at least 60%, 65%, 70%, 75% identical to each other remain hybridized to each other, whereas molecules with low percent identity cannot remain hybridized. A preferred, non-limiting example of highly stringent hybridization conditions are hybridization in 6 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C.

When hybridization occurs in an antiparallel configuration between two single-stranded nucleotides, the reaction is called "annealing" and those nucleotides are described as "complementary". A double-stranded nucleotide can be "complementary" or "homologous" to another nucleotide, if hybridization can occur between one of the strands of the first nucleotide and the second. "Complementarity" or "homology" (the degree that one nucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to hydrogen bond with each other, according to generally accepted base-pairing rules.

As used herein, the term "nucleic acid molecule" is intended to include DNA molecules, *e.g.*, cDNA or genomic DNA, and RNA molecules, *e.g.*, mRNA, and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules, which are separated from other nucleic acid molecules that are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated marker nucleic acid molecule of the invention, or nucleic acid molecule encoding a peptide marker of the invention, can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence as a hybridization probe, a molecule comprising a nucleotide sequence of the present invention can be isolated using standard hybridization and cloning techniques as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate nucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, nucleotides corresponding to marker nucleotide sequences, or nucleotide sequences encoding a marker of the invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

A nucleic acid molecule of the invention, moreover, can comprise only a portion of the nucleic acid sequence of the invention, or a fragment which can be used as a probe or primer. The probe/primer typically comprises substantially purified nucleotide.

Probes based on the nucleotide sequence of a nucleic acid molecule encoding a peptide of the present invention can be used to detect agglomeration proteins. In other embodiments, the probe comprises a labeling group attached thereto, *e.g.*, the labeling group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpresses, *e.g.*, over- or under-express, a polypeptide of the invention, or which have greater or fewer copies of a gene of the invention.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6 X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO. 1-10. As used herein, a "naturally-occurring" nucleic acid molecule includes an RNA or DNA molecule having a nucleotide sequence that occurs in nature, *e.g.*, encodes a natural protein.

In other embodiments, the nucleotides of the invention can include other appended groups such as peptides, *e.g.*, for targeting host cell receptors *in vivo*, or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W0 89/10134). In addition, nucleotides can be modified with hybridization-triggered cleavage agents (*see*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents (*see*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the nucleotide may be conjugated to another molecule, *e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent. Finally, the nucleotide may be detectably labeled, either such that the label is detected by the addition of another reagent, *e.g.*, a substrate for an enzymatic label, or is detectable immediately upon hybridization of the nucleotide, *e.g.*, a radioactive label or a fluorescent label, *e.g.*, a molecular beacon as described in U.S. Patent 5,876,930.

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a marker protein of the invention (or a portion thereof). As used herein, the term "vector" includes a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which includes a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced, *e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors. Other vectors, *e.g.*, non-episomal mammalian vectors, are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors." In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operatively linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence, *e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell. The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements, *e.g.*, polyadenylation signals. Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide

sequence only in certain host cells, *e.g.*, tissue-specific regulatory sequences. It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein, *e.g.*, marker proteins, mutant forms of marker proteins, fusion proteins, and the like.

The recombinant expression vectors of the invention can be designed for expression of marker proteins in prokaryotic or eukaryotic cells. For example, proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in marker activity assays, *e.g.*, direct assays or competitive assays described in detail below, or to generate antibodies specific for marker proteins for example.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

Another aspect of the invention pertains to host cells into which a nucleic acid molecule of the invention is introduced within a recombinant expression vector or a nucleic acid molecule of the invention containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur

in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. Preferably, the host cell is a prokaryotic cell. For example, the invention can be expressed in bacterial cells such as
E. coli. Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid, *e.g.*, DNA, into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

A host cell of the invention, such as a host cell in culture, can be used to produce, *i.e.*, express, a recombinant protein. Accordingly, the invention further provides methods for producing a protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a protein, or proteins, has been introduced) in a suitable medium such that a protein of the invention is produced. In another embodiment, the method further comprises isolating a protein from the medium or the host cell.

Of course, one skilled in the art will appreciate further features and advantages of the invention based on the above-described embodiments. Accordingly, the invention is not to be limited by what has been particularly shown and described, except as indicated by the appended claims.

EXAMPLES

Example 1: Detection of transient gene expression in single living *E.coli* cells with sensitivity for one protein molecule

(i) Construction of a short-lived β-gal

As discussed *supra*, in order to observe individual events involved in the expression of the *lacZ* gene, one must construct an *E. coli* strain that expresses short-lived *β*-gal. To achieve this goal, the inventors employed the so-called N-end rule (Tobias, J.W., *et al.*, Science, 1991. **254**(5036): p. 1374-7, and Varshavsky, A., Proc Natl Acad Sci USA, 1996. **93**(22): p. 12142-9, the entire teaching of which is incorporated herein by reference) and N-terminal signal peptides (Flynn, JM. *et al.*, Molecular Cell, 2003, 11, 671-683, the entire teaching of which is incorporated herein by reference) to shorten the cellular lifetime of *β*-gal. The N-end rule states that the cellular lifetime of a protein is related to its N-terminal amino acid residue. In *E. coli*, changing N-terminal amino acid from the natural methionine to leucine, arginine, lysine, phenylalanine, tryptophan or tyrosine shortens the protein's half-life to a few minutes. In this experiment, the ubiquitin fusion technique was used in order to introduce a lifetime-shortening amino acid (*e.g.*, leucine or arginine) replacing the methionine at the N-terminus of β-gal to generate Ub-Leu-*β*-gal or Ub-Arg-*β*-gal , see, Bachmair, A., D. Finley, and A. Varshavsky, Science, 1986. **234**(4773): p. 179-86, the entire teaching of which is incorporated herein by reference. After this fusion protein is expressed, the ubiquitin is cleaved by ubiquitin-specific protease, thus the argine or the leucine residue is exposed to the proteolytic pathways in *E.coli*.

An *ub-srg-lacZ* fusion gene was constructed on a plasmid. However, expressing the fusion gene off of the constructed plasmid would introduce many complications due to the variable copy number of plasmids from cell to cell. Therefore, in order to observe stochastic expression of *lacZ* at the basal level, this fusion gene was integrated into the *E. coli* genome by homologous recombination, see, Link, *et al.*, J Bacteriol, 1997. **179**(20): p. 6228-37, the entire teaching of which is incorporated herein by reference. The same method was used to construct a *lacZ*

strain, in which the entire coding sequence of β-gal is deleted. Hydrolysis of DDAO-gal in this *lacZ*⁻ strain is essentially abolished as compared to that of wild type (see, FIG. 4). This demonstrated that the hydrolysis of DDAO-Gal by β-gal is specific and the fluorescence signal observed is related to the expression of β-gal only.

After replacing the endogenous β-gal gene with Ub-Arg-β-gal, the plasmid pRB293, containing the UBP1 gene, the *Saccharomyces cerevisiae* ubiquitin-specific processing protease (see, Tobias, J.W. and A. Varshavsky, J Biol Chem, 1991. **266**(18): p.
12021-8, the entire teaching of which is incorporated herein by reference) was transformed into the cell. The altered cellular lifetime of Ub-Arg-β-gal using spectroscopic methods was then examined.

(ii) Live cell observation.

After obtaining the Ub-Arg-β-gal construct, live cell experiments were conducted employing a total internal reflection fluorescence (TIRF) microscope. The experiments were conducted using various concentrations from 1 μM to 50 μM of DDAO-gal in M9 minimal media, which is perfused through the sample chamber above the *E. coli* cells pushed down on the glass coverslip by a droplet of agarose gel. Detectable DDAO fluorescence bursts associated with the stochastic events of gene expression even at the reduced number of β-gal were observed (see, FIG. 7). Most . importantly, the time traces of the fluorescence bursts exhibit quantized levels corresponding to β-gal molecules generated and degraded one at a time. This demonstrated that this method has the ultimate sensitivity for even one protein molecule. From these individual *lacZ* expression events, important parameters were extracted, such as the dissociation constant k_d between the *lac* repressor and the operator and the expression efficiency in the cellular environment. These measurements are important because the thermodynamics and kinetics of biochemical reactions, in principle, can be distinctly different in the cellular environment than *in vitro*. The previous understanding of these parameters were either obtained from *in vitro* experiments or deduced indirectly from ensemble cellular measurements.

(iii) Expanding reporter repertoire.

β-gal is but one system to demonstrate the proof of principle of short-lived reporter proteins. This same strategy described herein can be used with other reporter genes in order to track transient behavior. These reporter genes can be used to make fusion proteins for multiplexing observation of gene expression processes. One will be able to study gene regulatory circuits by examining the effects of one gene on another. Such work will offer detailed information about the interactions and regulation among gene products. For example, another reporter, β-glucosidase with a molecular mass of 82 kDa, encoded by the gene *bgl*B from Bacillus sp. GL1 (Arch. Ciochem. Biophys., vol 360, No. 1, pp 1-9, 1998) is employed. This enzyme hydrolyzes the non-reducing terminal glucoside from either carbon hydrates or artificial substrates such as resorufin-glucopyranoside (see FIG. 1 (c) and (d) for the substrate structure and product spectrum). We have expressed recombinant β-D-glucosidase in *E. Coli*. The strain that express the β-glucosidase gene showed very high hydrolysis activity on resorufin-glucopyranoside, while the wild type *E. Coli* (does not contain the gene encoding for β-glucosidase) showed negligible glucosidase activity (FIG. 9). As another example, β-lactamase, which hydrolyzes fluoregenic substrates such as CCF2 and CR2/AM (see, Zlokarnik *et al.*, Science, 1998, 279(5347), 84-88, Gao *et al.*, J.Am.Chem. Soc, 2003, 125, 11146-11147, the entire teachings of which are incorporated herein by reference.), can also be genetically modified and employed in the reporting system.

Example 2: Construction of live cell array of *E.coli*

The construction of two libraries comprised of short-lived β -gal and short-lived YFP and the corresponding live cell arrays are illustrated in *E.coli* as described in detail in the following. However, it should be obvious to those skilled in the art that other short-lived reporter genes such as β-glucosidase or β -lactamase , and other organisms such as *Saccharomyce cerevisiae* and *Shewanella oneideinsis* can equally be the subjects of the present invention.

(i) Construction of a short-lived β-gal and YFP reporter libraries of *E.coli*

Investigators will proceed with an *in vitro* Tn5-based transposon system, (see, Goryshin, I.Y., et al., Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol, 2000. **18**(1): p. 97-100, the entire
teaching of which is incorporated herein by reference)

To generate the *lacZ* library using the *in vitro* Tn5 mediated transposition, a DNA cassette including a promoter-less *ub*-x-*lacZ* gene (the *x* between *ub* and *lacZ* represents any amino acid that shortens the cellular lifetime of the resulting β-gal) will be cloned into a transposon construction vector pMOD-2 (Epicentre Technologies), flanked by two Tn5-recognizable 19 bp ME sequence. This *ub*-x-*lacZ* gene contains its own ribosome binding site (RBS), in front of which a stop codon will be placed to avoid a translation read-through from a previous gene. See, FIG. 10.

Figure 10 is a schematic drawing of the construction of the *lacZ* library by Tn5 mediated transposition. ME represents Tn5 recognizable mosaic ends sequence (triangles); RBS are the ribosome binding sites (rectangles); and the box joined by a hitched box indicates the *ub*-x-*lacZ* gene and the oval with a turn arrow on top indicates a promoter on the chromosome.

The selection for desired colonies containing the reporter genes will be based on blue/white colony screening on X-gal plates. The expression of the promoter-less *Ub-X-lacZ* gene from a functional promoter on the chromosome will result in blue colonies due to the conversion of X-gal into blue insoluble precipitant by β-gal. Since the conversion of X-gal by β-gal is highly efficient and can accumulate, even colonies transiently expressing β-gal can be identified if sufficient growing time is allowed. Investigators have observed that the *E. coli* colonies contain one single copy or less of short-lived β-gal on average produce easily visible blue color after 16 hours incubation. Thus, investigators expect nearly all promoters, even the tightly controlled promoters at its basal level activity can be identified using this blue/white screening.

Figure 11 depicts an alternative method for constructing the *lacZ* and *YFP* libraries simultaneously. In this method, the last selection step generates both *lacZ* and *yfp* libraries based on blue and white colonies screening. (The notations are the same as FIG. 10.)

The approach for constructing the two libraries simultaneously is planned as follows: First, a methylated DNA cassette will be randomly inserted into *E.coli* genome by Tn5 mediated *in vitro* transposition as described above. This DNA cassette will contain a copy of *ub*-x-*lacZ* (contains a stop codon and its own ribosome binding site in front), and also a copy of Venus-ssrA with its 3' end flanked by approximately 500 bp sequence, which is homologous to the 3' end of the *lacZ* gene. Between the *ub*-x-*lacZ* and Venus-ssrA lies the *cat* and *sacB* genes. The first round selection for the incorporation of this DNA cassette into the chromosome will be based on the β-galactosidase activity on the X-gal plate or chloramphenicol resistance. The colonies from the first round of selection will be pooled and plated on sucrose plates supplied with X-gal. Blue colonies that survived on the sucrose plates indicate the presence of the *ub*-x-*lacZ* gene on the chromosome, thus forming the *lacZ* library, while the white colonies indicate the presence of Venus-ssrA, forming the YFP library. Both libraries will then be replicated on chloramphenicol plates to ensure that the survival on sucrose plates is not due to the mutation of the *sacB* gene (see, Link, A.J., D. Phillips, and G.M. Church).

The investigators are also aware that the construction of the two libraries will probably still leave some promoters not covered. In such cases, insertion of the reporter genes after the specific promoters will be done separately using homologous recombination individually, assuming the numbers of the promoters not covered by the methods described above is not significant.

All the above work can be automated as explained in the following: the blue colony picking, inoculation into 96-well plates, and the subsequent master plates making with arranged colonies will be performed by the Q-bot (Genetix). See, Fig. 12. Colonies from the master plates will be picked directly into 96-well PCR tubes containing reaction buffers prepared by the Genesis liquid-handling robot (Tecan).

The PCR reactions will then be cleaned using Qiaquick 96 PCR purification kit (Qiagen Inc.) on a Beckman BioMek FX robot. Automated DNA sequencing will be performed by a commercial company. The process of reading the DNA sequence files, blasting and mapping onto the *Shewanella* genome can also be automated by a home-made program.

To identify the position of the reporter gene on the chromosome, regions before and after the Ub-x-*lacZ* gene in each strain will be sequenced. A modified colony PCR using a protocol called random amplification of transposon ends (RATE), (see, Karlyshev, A.V., M.J. Pallen, and B.W. Wren, Single-primer PCR procedure for rapid identification of transposon insertion sites. Biotechniques, 2000. **28**(6): p. 1078, 1080, 1082, the entire teaching of which is incorporated herein by reference), will be employed to amplify the regions around the *ub*-x-*lacZ* gene. Abundant single-stranded DNA (ssDNA) will be first generated by one primer, which specifically targets one end of the *ub*-x-*lacZ* gene and goes outward relative to the transposon DNA. Second, these ssDNAs will be amplified by random priming at low annealing temperature using the same primer to produce double-stranded DNA (dsDNA) with different lengths. Finally, these dsDNAs will be used as templates and amplified using the same primer at stringent annealing temperature. A new primer, which targets specifically a sequence lying in the middle of the ME sequence and the first primer binding sequence on the transposon, will be used to sequence the amplified dsDNA. The sequence will then be compared to *E.coli* genome to identify the position of insertion.

At this step, investigators will pick at least 2×10^4 blue colonies to establish the initial library. At this size, one would expect on average one insertion per 250 bp on the genome (the genome size of *E.coli* is approximately 4.6×10^6). The goal is to tag every possible promoter (few thousands in total, judging by the predicted 4000 genes in *E.coli*) with a reporter. Therefore, the activity of each promoter or operon of the entire genome in response to different stimuli can be studied in parallel on one or two live cell arrays. To achieve this goal, investigators will select from the initial library based on the sequence data according to the following criteria: 1) least

disruption of a gene; 2) least polar effect to the downstream genes caused by the insertion; and 3) at least one insertion for one promoter or operon.

The final library is estimated to contain at least 3000 strains including both unique and multiple reporter gene insertions for each predicted promoter.

(b) Construction of live-cell microarrays

Once the construction of the reporter library is completed, investigators will print the reporter-labeled cell strains into microarrays. Nanolitres of aqueous media containing *E.coli* cells will be pipetted onto a #1 glass coverslip using a robotic micro-arrayer; the Omnigrid (GeneMachine) arrayer is capable of dispensing a minimum of 300 picolitre of fluid. Sub-microlitres of low-melting-temperature agarose containing growth media will be immediately applied on top of the cell solutions to prevent drying of cells. The weight of the agarose will compress the media droplets and create a monolayer of cells on the surface of the cover glass. The resulting spot will be approximately hundreds of micrometers in diameter, which is about the size of the view-field on a microscope. Macroscopic version of this technique has been consistently demonstrated, and cells stay viable and divide for many generations on the slide.

Allowing the spacing between adjacent spots to be 1mm, it is possible to print an array of micro-colonies corresponding to the library of 1000 strains in a 60mm x 60mm area. That is well within the travel range of an automated x-y stage.

Once the array is printed, it will be capped by a gasket and Microaqueduct slide manufactured by Bioptech Inc, as shown in FIG. 12. The microaqueduct slide will allow laminar flow through the chamber and keep the temperature constant via an add-on thermoelectric heater unit. A solution of DDAO-gal and growth media can be perfused through the chamber to keep the cells supplied with nutrients and fluorogenic substrates, while allowing fluorescent product DDAO and cellular metabolites to flow away. Unlike liquid suspensions, this set-up, whereby media is allowed to flow over microbes that are fixed in place, is very similar to those

routinely used for analysis of biofilm formation, and is therefore, more representative of how bacteria exist in the environment.

The microarray encased in a flow chamber offers a versatile and durable platform with a controlled environment and constant supply of nutrients. This chamber will be mounted on TIRFM microscope (Nikon Te-2000E) with a built-in motorized XY stage. Equipped with rotary encoders and feedback stepper motors, the XY stage can visit each micro-colony on the microarray with a repeatability of one micron. With an external Z-drive, the objective lens can auto-focus before acquiring an image at each spot. Shutters and filter wheels controlled by commercial software (such as Metamorph, Universal Imagining Inc.) can precisely time illumination with laser and Xe lamp, to acquire fluorescent and phase-contrast images. These images can be stacked into movies for each point in the microarray. Fluorescent time trajectories will be extracted for each individual cell and proper statistics analysis will be performed.

Investigators will conduct experiments to monitor gene expression responses to various environmental factors with a complete library. Since the environmental changes and factors are uniform for all microcolonies on the array, observation of fluorescence changes at each spot will reflect changes in expression of each tagged promoter induced by the stimuli. Given the capability of the motorized stage and the control software, it can easily scan 100 spots per minute, collecting time-point trajectories of approximately 100 cells at each spot.

High-throughput real-time data provides quantitative information on system-wide gene expression kinetics. This first-of-a-kind systems biology dataset provides an opportunity for mathematical modeling.

Example 3: Real-time gene expression of live *Shewanella oneideinsis*

(i) Demonstration of DDAO-gal permeability

Investigators have demonstrated that the fluorogenic substrate of β-gal, DDAO-gal, is permeable to the *Shewanella oneidensis* cell membrane. To do so, they have transformed wild-type (*lacZ*) *Shewanella oneidensis* cells with pBBR1MCS5.1 (see, FIG. 13), a plasmid containing the *lac* operon along with the *lacI* repressor gene. Figure 10a depicts a plasmid map of the pBBR1MCS-5.1 plasmid. Figure 13b is the nucleotide sequence [SEQ ID NO. 26] encoding the ubiquitin and part of the beginning of β-gal on the plasmid pBBR1MCS5.1.

Figure 14a shows the fluorescence image of the individual transformed cells supplied with DDAO-gal without induction. In contrast, under the same condition, no fluorescence signal was observed in the wild-type strain, see, FIG. 14b. This experiment proves that DDAO-gal can permeate through the *Shewanella oneideinsis* cell membrane and the fluorescence signal is specifically due to the presence of β-gal.

(ii) Demonstration of a short-lived X-β-gal in *Shewanella*

The N-end rule has been demonstrated to be universal in organisms examined such as *E. coli*, yeast and mammals, (see, Varshavsky, A., The N-end rule: functions, mysteries, uses. Proc Natl Acad Sci U S A, 1996. **93**(22): p. 12142-9, the entire teaching of which is incorporated herein by reference). *Shewanella* is closely related to *E. coli*, therefore, it is reasonable to assume that the same rule also applies in *Shewanella*. As shown in FIG. 14b, when a short-lived β-gal (Ub-Leu-β-gal) is expressed together with the ubiquitin-specific protease, the hydrolysis rate decrease dramatically, indicating shortened cellular lifetime of β-gal.

Example 3: β-gal applied to *Saccharomyce cerevisiae*

A short-lived version of β-gal (*ub-leu-lacZ*) was used in the eukaryotic model organism *Saccharomyce cerevisiae* (budding yeast) to probe stochastic gene expression events. *Saccharyomyce cerevisiae* has extensive ubiquitin-dependent protein degradation pathways, thereby enabling a cellular lifetime of modified β-gal less than a few minutes.

The *ub-leu-lacZ* reporter gene was generated using standard cloning protocols (Sambrook and Russell, Molecular Cloning, 3rd Ed, CSHL press, the entire teaching

of which is incorporated herein by reference) with a pair of PCR primers (5' CTTGGTA CCATGCAGATTTTCGTCAAGACTTTG 3' [SEQ ID NO. 27], and 5' GAGCGGC CGCTTTTGACACCAGACC 3' [SEQ ID NO. 28]) to amplify a 4000bp fragment containing *ub-leu-lacZ* from pUB23 plasmid generated by Varshavsky, *et al.* This DNA fragment was ligated into the pYC2/CT plasmid (Invitrogen, Inc.) and the resulting construct was verified by DNA sequencing.

Figure 15a depicts the nucleotide sequence junction of *ub-leu-lacZ*, and (b) is the amino acid sequence [SEQ ID NO. 29] and nucleic acid sequence [SEQ ID NO. 30] for the junction of the *ub-leu-lacZ* construct on centromeric plasmid: the sequence is from the Gal1 promoter site to the Bsu26I site of the *lacZ* gene, and numbering of the nucleotides is according to the first base of the Gal1 promoter; the ubiquitin gene is joined by an modified *lacZ* gene with its first methionine residue replaced by a leucine residue. The amino acids sequences are shown on top of the DNA sequence panel.

Figure 16 shows DDAO fluorescence generated from the hydrolysis of DDAO-gal by $lacZ^+$ (dark) but not by the $lacZ^-$ (light) yeast cells measured in a fluorometer. Final concentration of DDAO-gal was 50 μM and *S. cerevisiae* cells was grown to middle log phase in synthetic dextrose medium. The significantly different hydrolysis rates between the two strains demonstrated that (i) fluorescence substrate DDAO-gal is permeable to *S. cerevisiae* cell wall and plasma membrane; (ii) DDAO-gal is hydrolyzed by β-gal with remarkable specificity and high turnover rate.

Figure 17 is a fluorescence image of *S. cerevisiae* cells expressing wild type β-gal. This experiment was done using fluorescence microscope with excitation at 568nm. The other setup is identical to those used in the *E.coli* and *Shewanella* experiments as described above. The presence of glucose in the growth media represses the *Gal1* promoter, resulting in a low basal level expression of β-gal. Experimental conditions were chosen to minimize the background and autofluorescence of yeast cells.

Figure 18 shows the fluorescence burst observed on a single *S. cerevisiae* cell with a short-lived β-gal expressed from the centromeric plasmid. The burst in the time trace indicates a single *lacZ* gene expression event, resulting from the stochastic dissociation of the repressor from its binding site. The rise of the burst indicates the generation of β-gal and the decay indicates the degradation of β-gal. This is the first experiment demonstrating that the short-lived β-gal reporter system is capable of detecting low copy number translational product and following the gene expression events in realtime in live eukaryotic cells.